DOCUMENT RESUME

ED 464 123 TM 033 804

AUTHOR O'Neill, Kathleen; Rizavi, Saba
TITLE Performance of Examinee Groups on a Measure of Analytical Writing.
PUB DATE 2002-04-00
NOTE 29p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *College Students; Ethnic Groups; Graduate Study; Higher Education; Racial Differences; Sex Differences; *Test Construction; *Writing Tests
IDENTIFIERS *Analytical Writing; *Graduate Record Examinations

ABSTRACT

Beginning in 1999, an analytical writing measure was offered as an optional test in the Graduate Record Examinations (GRE) program. This test will be incorporated into the GRE General Test in fall 2002. The essays in this test focus on critical reasoning and analytical writing skills. This study examined the performance of various examinee groups on this performance-based test. Because the test has been optional, the examinees are self-selected in that they presumably thought that they would do well on the analytical writing tasks. Over the 2-year period, 5,946 examinee records could be matched to both the Writing and GRE General Test. Results provide some assurance that incorporating a writing measure into the GRE General Test will not necessarily increase group differences. Standardized differences for gender groups and many minority groups in this study are smaller on the Writing Assessment than on the multiple-choice Analytical measure, which the Writing Assessment is to replace. For some groups, most notably examinees who may have second language difficulties such as English-as-a-Second-Language and Asian American examinees, the standardized differences are larger on Writing than on Analytical. From a fairness perspective, these results have implications for scoring the assessment. The decision to use the writing measure as part of an admissions decision could have a major effect on applicants selected for admission. Allowing a choice of Issue topics appears to be associated with differential preference rates some of the time. It is also apparent that the time allotment for essay tasks has major consequences for some examinee groups. Five appendixes contain supplemental material in tables. (Contains 13 tables, 12 figures, and 6 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 123

Performance of Examinee Groups on A Measure of Analytical Writing

Kathleen O'Neill and Saba Rizavi, ETS

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

K. ONeill

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

National Council on Measurement in Education

New Orleans, LA

April 2002

TM033804

BEST COPY AVAILABLE

Performance of Examinee Groups on A Measure of Analytical Writing

Though admissions tests for graduate school have long used multiple-choice questions, they have far less often included performance assessments. Increasingly, faculty making graduate admissions decisions have wanted to know more about the candidates' ability to think critically and to express themselves cogently – skills seen as essential in most graduate departments. To this end, an analytical writing measure was offered as an optional test in the Graduate Record Examinations (GRE) program beginning in 1999. (This test will be included in the GRE General Test beginning in Fall 2002.) The essays in this test focus on critical reasoning and analytical writing skills such as the ability to articulate and support complex ideas, to analyze an argument, and to sustain a focused and critical discussion. This study will report on the performance of various examinee groups on this performance-based test.

Objectives of the Investigation

The purposes of the investigation were to examine:

a) the performance of various examinee subgroups. The primary focus for this topic was to determine whether differential performance existed and, if so, to quantify the amount of the difference. The groups of interest included male and female examinees, minority and White examinees, ESL examinees, and examinees in different age groups and in various intended major fields (e.g., Education, Engineering, Natural Science).

b) the amount of differential speededness for examinee subgroups, if any.

c) differential selection in the choice of an Issue topic.

d) the comparative difficulty of different prompt content areas.

Based on findings from other tests (e.g., Breland, Bridgeman, and Fowles, 1999; Breland, Muraki, and Lee, in preparation; Bridgeman and McHale, 1996; Willingham and Cole, 1997), the GRE program hypothesized some findings – for example, that gender differences would be minimal. However, there were other areas of interest for which hypotheses were less clear – such as the performance of groups by intended major field. In addition, there was interest in the performance of prompts (e.g., topic chosen and time used).

Description of the Instrument

The GRE Writing Assessment, which began operational testing on October 1, 1999 as a standalone test, was developed to meet the needs of graduate programs for an assessment of critical thinking and analytical writing skills. The test contains:

1) a 45-minute task entitled "Present Your Perspective on an Issue" in which examinees state their opinion on an issue of general interest from any perspective and provide relevant reasons and examples to explain and support their view. Examinees may choose one of two topics presented to them. Choice was offered because the task requires examinees to develop the content for their response by drawing upon their own reading, observation, and/or experience.

2) a 30-minute task entitled "Analyze an Argument" in which they critique an argument presented to them by discussing how well reasoned it is and considering the logical soundness of the argument.

These two essay tasks are complementary in that one requires the examinee to construct his/her own argument about an issue, whereas the other requires the examinee to critique someone else's argument by assessing its claims and evaluating the evidence it provides. All 125 essay topics for each task type are published on the Web in order to facilitate examinee preparation.

When taking the test, examinees may choose to word process or handwrite their essays. Each task is holistically scored by two readers on a six-point scale,[1] and the scoring rubric is based on the quality of writing demonstrated in the two tasks. Because the assessment measures analytical writing, skills such as the ability to reason, to assemble evidence to develop a position, and to communicate complex ideas weigh more heavily in scoring than does the writer's control of fine points of grammar or the mechanics of writing (e.g., spelling). A single test score is reported on a six-point scale, with half-point intervals. The single test score is calculated by averaging the two readers' scores for each task type and then averaging the two task type scores. The final score is rounded up to the nearest half point (e.g., 4.25 is rounded to 4.5).

Information about the reliability of the test and correlations with the General Test is contained in Appendix A, and a fuller description of the psychometric qualities of the test is contained in Shaeffer, Briel, and Fowles (2001), available on the GRE Web site (www.gre.org).

Description of Sample

Because the writing measure is relatively new to the GRE program and because there is a time lag for departments to decide to require a new test, the examinees in these

analyses are, for the most part, self-selected, and they are not submitting their score to meet a graduate school requirement. Presumably, the examinees felt that they had the ability to do well on the analytical writing tasks. Though they may be self-selected, there is no reason to anticipate that there would be major differences in various gender, language, or ethnic groups as a result (e.g., both men and women presumably felt that they had sufficient ability to do well on the analytical writing tasks).

The effect of the self-selection is most obvious in the distribution of the examinees within major fields, as shown in the chart below.

Major Field Group	% in GRE Population	% in Writing Population[2]
Business	2	3
Education	9	7
Engineering	11	2
Humanities	10	17
Natural Sciences	37	35
Social Sciences	19	22
Other	12	12

Humanities majors are over-represented in the Writing Assessment population compared to the typical GRE population and Engineering majors are underrepresented.

The total number of examinees who responded to both topics in the Writing Assessment over the two-year examination period was 8736. Of these, only 5946 examinee records (68%) could be matched to the GRE General Test. Examinees who also took the General Test performed better on the Writing Assessment than those who did not; it may be that these examinees were more serious about their applications for graduate study than the unmatched group.

---

[1] About 60% of the time the two scorers agree exactly. About 98% of the time they agree within one point.
[2] On the Writing Assessment, 9% of examinees who provided demographic information did not indicate their major field.

Analyses

Means, standard deviations, and standardized differences on the Writing Assessment were calculated for gender, minority, language group and major field groups. Analyses of variance were used to test significance of differences between means. Standardized differences on V, Q, and A were also calculated for the subset of examinees who took the General Test.

The choice of topics was evaluated for examinees at three different ability levels within groups. For these analyses, topics were clustered based on their classifications – which involved aspects of content and the types of reasoning strategies employed.

Results

Total Group

The mean total score on the Writing Assessment for the total group was 4.27 (SD = 0.90). Examinees who also took the GRE General Test (called matched examinees) performed better than examinees who did not, as shown below.

	N	Mean	SD
Total group	8736	4.27	0.90
Those with GRE General Test scores	5946	4.35	0.88
Those without GRE General Test scores	2790	4.08	0.92

For the 5,946 examinees who had both Writing and GRE General Test scores, the General Test scores are shown below; these are contrasted with the scores for the GRE norm group.[3]

[3] The norms group contains all examinees who tested between 10/1/97 and 9/30/2000. – more than 1,000,000 for each test.

Test	Writing Examinees with General Test Scores		Examinees in GRE Norm Group	
	Mean	SD	Mean	SD
Verbal	476	114	470	115
Quantitative	540	134	576	145
Analytical	560	141	552	135

In comparison with the whole GRE General Test population, Writing Assessment examinees with General Test scores perform about as well on the Verbal and Analytical measures and slightly less well on the Quantitative measure.

Gender, Language, and Minority Groups

Mean scores on the Writing Assessment for various examinee groups on the two task types are shown below.

Group	N	Issue Mean (SD)	Argument Mean (SD)
Gender:			
Females	4053	4.37 (0.88)	4.10 (1.03)
Males	1736	4.41 (0.99)	4.10 (1.09)
Language:[4]			
ESL	479	4.23 (0.96)	4.00 (1.08)
Non ESL	5169	4.38 (0.86)	4.11 (1.04)
Minority:			
American Indian	33	4.50 (0.93)	3.98 (1.20)
African American	424	3.90 (0.93)	3.44 (1.05)
Mexican American	99	4.23 (0.84)	3.94 (0.91)
Asian American/ Pacific Islander	284	4.04 (0.98)	3.92 (1.16)
Puerto Rican	52	3.76 (1.14)	3.46 (1.11)
Other Hispanic	110	4.22 (0.89)	3.75 (1.05)
Other	167	4.46 (0.92)	4.23 (1.08)
White	4051	4.50 (0.85)	4.23 (0.99)

These results show that, for each group, examinees score better on the Issue task than they do on the Argument task, a finding that may be partly due to the nature of the task,

to the choice available on this task, and/or to the time available for the task. For most groups, the standard deviation is smaller for the Issue task than for the Argument task. (Score distributions for various examinee subgroups are shown in Appendix B.)

In the gender group comparisons, women are scoring about the same as men on both tasks; this similar performance contrasts with the multiple-choice questions in the Verbal measure on which men outscore women. As might be predicted, ESL students perform less well than do non ESL students, but the difference between the two groups is not as large as might have been anticipated. The difference in means for the two groups is statistically significant (at the .05 level) only for the Issue task.

For minority groups, there are smaller differences on the Writing Assessment than on the Verbal or Analytical measures of the General Test. (The exception: a slightly greater difference for Asian American examinees on the Writing Assessment than on the Verbal or Analytical measures.) Nonetheless, it is clear that the means for African American, Asian American, and White examinees differ significantly (at the .05 level) for both essay task types.

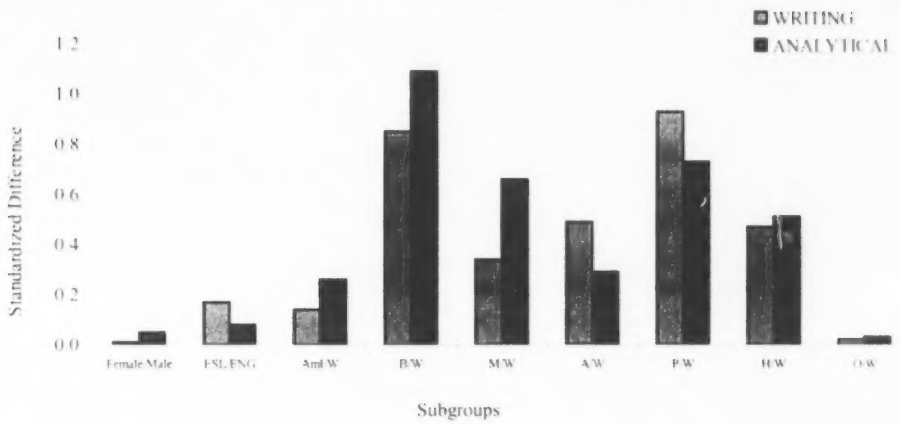
In the figure below are shown standardized differences[5] for matched examinees on the Writing Assessment and on the Analytical measure (which will be replaced by Writing in October 2002). (Standardized differences for the other General Test measures are shown in Appendix C.)

---

[4]Examinees were considered to be ESL examinees if they answered "No" to the question: Do you communicate better (or as well) in English than in any other language?

[5] The standardized difference is computed by subtracting the female mean score from the male mean over the pooled standard deviation and by subtracting the ESL mean from the English-best-language mean over the pooled standard deviation. For the ethnic groups, each ethnic group is subtracted from the White mean.

Standardized Differences between Subgroups



These results show that the Writing Assessment had much smaller group differences than the Analytical measure for women, African American, and Mexican American examinees. However, group differences were larger for ESL examinees, and Asian American and Puerto Rican examinees (though the sample size in the latter group is small enough to warrant caution about these results).

Age Groups

The performance of examinees in various age groups was investigated because there was a hypothesis that older examinees, who are more likely to have work experience entailing writing, might perform better than examinees who proceed straight from undergraduate to graduate work. The performance of examinees by age group is shown below.

Age Group[6]	N	Issue Mean (SD)	Argument Mean (SD)
18-20	17	4.62 (0.99)	4.59 (0.91)
20-24	3883	4.43 (0.86)	4.17 (1.01)
25-29	2197	4.19 (0.94)	3.94 (1.06)
30-34	985	4.19 (1.00)	3.92 (1.12)
35-39	561	4.10 (1.01)	3.79 (1.17)
40-44	449	4.05 (1.09)	3.67 (1.12)
45-49	345	4.17 (0.96)	3.78 (1.10)
50-54	212	4.20 (1.05)	3.81 (1.13)
55-59	59	4.29 (1.00)	3.89 (1.06)
60-71	17	3.91 (0.97)	3.47 (0.98)

These results do not support such a hypothesis; in fact, examinees in the most traditional age group (20-24) scored higher than all older groups on both task types for this self-selected population. One factor clouding the interpretation above is that age is likely to be confounded with major field: there are more older Education majors (a lower scoring group) than there are Natural Science majors (a higher scoring group).

Major Field Groups

Examinees in some major fields might be expected to perform particularly well on the Writing Assessment because their curricula provides more emphasis on writing. Mean scores for examinees in various major fields are shown below.

[6] Because the age groups are clustered based on groups of interest, the interval sizes are not equal.

Major Field Group	N	Writing Mean (SD)
Business	187	4.20 (0.86)
Education	403	4.05 (0.87)
Engineering	131	4.11 (0.99)
Humanities	933	4.66 (.087)
Natural Sciences	1896	4.28 (0.83)
Social Sciences	1191	4.48 (0.84)
Other	651	4.40 (0.87)

Not surprisingly, examinees intending to major in Humanities areas perform better than other examinees, as do Social Science majors. Examinees from Engineering and Education fields perform worse than other examinees. These differences may reflect, in part, the differences in curricula: Humanities and Social Science majors may be more likely to write papers and essay examinations in their undergraduate coursework than other majors.

Testing Time

A differential use of testing time may provide insight into differences in group performance. For all groups on the Writing Assessment, more examinees use the full amount of time for the Argument task than for the Issue task. One obvious explanation for this difference is that examinees are allowed more time for the Issue task. Another hypothesis for this finding is that examinees are spending any leftover time reviewing the logic of the argument even though they have constructed their essay. It could also be that examinees are working more slowly on their second essay task. However, even when examinees are spending the full amount of time on essays, scorers have noticed no evidence that examinees have left ideas unstated (e.g., by ending in mid-sentence, or by omitting any sort of concluding statement). Score distributions for the essays also do not

show any skew toward lower scores (which could indicate incomplete thoughts). Moreover, in a psychometric study conducted while the Writing Assessment was in development, students reported that they felt that they had sufficient time to complete the tasks (Shaeffer, Briel, and Fowles, 2001).

Figures showing the distributions of subgroup completion times for each writing task are contained in Appendix D and Chi-square results are shown in Appendix E. In general, men are slightly more likely to spend the full amount of time than women, and ESL examinees are slightly more likely to spend the full amount of time on the Argument task. Examinees from different minority groups show more differences in time utilization, as shown below.

Group	N	% Completing Issue Task		
		< 40 min	40 to < 44 min	44 – 45 min
American Indian	33	12	67	21
African American	424	26	59	15
Mexican American	99	20	72	8
Asian American	284	14	66	20
Puerto Rican	52	17	72	12
Other Hispanic	110	25	63	13
White	4051	23	62	15
Other	167	21	60	19

Group	N	% Completing Argument Task		
		< 25 min	25 – < 29 min	29 – 30 min
American Indian	33	12	18	70
African American	424	21	13	66
Mexican American	99	16	9	75
Asian American	284	7	8	85
Puerto Rican	52	17	13	69
Other Hispanic	110	21	12	67
White	4051	18	16	66
Other	167	15	13	72

These tables show that proportionally more African American examinees exit the Issue and the Argument tasks with 5 minutes remaining than do examinees in other groups[7]. For the Issue prompt, fewer Mexican American examinees use the full amount of time compared to other groups. For the Argument prompt, many more Asian American examinees are using the full amount of time compared to other groups.

When examinees are sorted by test score into three groups, the high scoring group is slightly more likely to spend the full testing time on each writing task than middle or low scoring examinees (see Appendix D).

Choice of Prompts

On the Issue task, examinees choose between two prompts that are presented. Topics for the pair to be selected are randomly selected after overlap issues (e.g. similar content) and difficulty have been taken into account. For the Issue prompts, there are ability level differences in the choices made by examinees for some of the Issue topics. The selection rate for 10 frequently chosen prompts is shown below.

Topic	Percent of Examinees Selecting the Topic		
	Low scorers	Middle scorers	High scorers
1	63	66	56
2	54	44	56
3	88	67	55
4	77	64	51
5	67	68	73
6	74	58	53
7	85	59	57
8	52	58	54
9	65	72	75
10	34	40	42

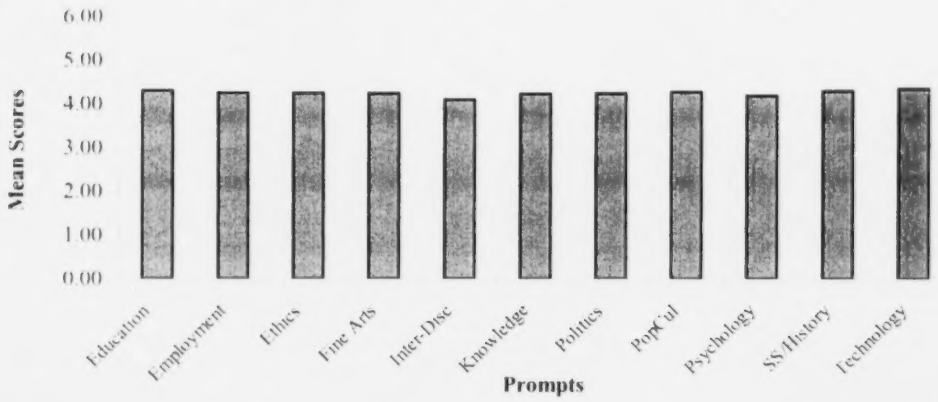
[7] Other Hispanic examinees also exit the Argument task with 5 minutes remaining at the same rate as African American examinees.

These results show that some topics (3, 4, 6, and 7) are chosen more frequently by low-scoring examinees than middle- and high-scoring examinees.[8] On other topics (5, 8, 10), the selection rate doesn't differ appreciably across the three ability levels. Investigations into the most popular prompts did not show a preponderance of topics from any single content area; each content area had several frequently chosen prompts.

Performance of Prompts

The performance of prompts based on different topic areas was investigated to see whether there were differences related to content, as shown in the figure below.

Mean Scores by Prompt Classfications



These results show there are no particularly difficult or particularly easy content areas.

Discussion

The results of this study provide some assurance that incorporating a writing measure into the GRE General Test will not necessarily increase group differences. Standardized

[8] It should be noted that examinees are sorted by scores, but the scores are necessarily affected by the prompt that was selected.

differences for gender groups and many minority groups in this study are smaller on the Writing Assessment than on the multiple-choice Analytical measure, which the Writing Assessment replaces. But for some groups, most notably examinees who may have second language difficulties such as ESL and Asian American examinees, the standardized differences are larger on Writing than on Analytical. From a fairness perspective, these results have implications for how the assessment is scored: scorers are specifically trained to ignore mechanics (e.g., spelling, punctuation) unless such features interfere with understanding. In addition, scorers are trained to accept any type of structure to an essay that meets the requirements of the prompt.

The performance of major field groups in this study shows that some major fields score lower than others. However, in multiple surveys by the GRE program, faculty in all fields have reported that writing is important in their discipline. The results of this study might then be due to the fact that writers better at analytical writing choose fields that emphasize analytical writing skills and/or that some disciplines require more analytical writing tasks and, thus, provide more practice in writing such essays.

One major concern for any test is the issue of speededness – especially differential speededness. Clearly, more examinees are using up the full testing time on the Argument task than on the Issue task, a finding that is related to the nature of the task and the time allotted for it. The pattern of time usage is variable across groups as well as tasks. Gender groups and language groups have roughly similar time usage patterns. Ability groups and minority groups show somewhat more variability.

Although there does not appear to be a relationship between difficulty and content area for this test, there clearly is a relationship between difficulty and task type. Examinees

are scoring higher on the Issue task. This result is not surprising since the task was structured to permit examinees to draw on their own backgrounds and interests and to permit choice in the topic. The incorporation of choice for the Issue task does add another factor into the way that examinees perform: for some prompts there are clear differences in choice levels for the low ability examinees.

These results lead to several major implications to this study. First, the decision to use a writing measure as part of an admissions decision could have a major effect on applicants selected for admissions. For example, in a field such as Physics which has a large number of foreign applicants, the use of the writing test in admissions decisions could result in fewer ESL students but more women and minority applicants. Even when the decisions about admissions might not be changed, there could be other ramifications. For example, despite recommendations from the GRE Program not to do so, some graduate schools award university-wide scholarships and fellowships based, in part, on GRE scores. To the extent that applicants are compared across academic fields, the Writing Assessment information could impact such awards. In such a system, English majors might see an increase in their chances of receiving such fellowships while Engineering majors might see a decrease.

Second, the issue of allowing choice of Issue topics appears to be associated with differential preference rates some of the time (a finding not unlike that for advanced placement tests – see Bridgeman, Morgan, & Wang, 1997). For some prompts, the low ability group was more likely to select the prompt than the higher ability groups. Although scorers are trained to adhere to a scoring rubric based on the full spectrum of possible scores, a concentration of low scorers in some prompts has implications for

researchers and developers. Further investigations of prompt choice will be conducted when the prompts are used with the complete GRE population, rather than a small self-selected one.

Third, the time allotment for essay tasks has major consequences for some examinee groups. It is clear that some examines are not using the full time allotment; when this time use is also associated with lower scores, examinees may be penalizing themselves. This has clear implications for the test preparation advice given to examinees, and it has ramifications that affect the validity of the test.

Bibliography

Breland, H. M., Bridgeman, B., and Fowles, M.E. (1999) Writing Assessment in Admission to Higher Education: Review and Framework. GRE Research Report 96-12R. NY: College Entrance Examination Board.

Breland, H., Muraki, E., & Lee, Y. (In preparation) Comparability of TOEFL CBT writing prompts for different prompt assignment and gender groups. Princeton, NJ: ETS.

Bridgeman, B. & McHale, F. (1996) Gender and Ethnic Group Differences on the GMAT Analytical Writing Assessment. ETS Research Report 96-2. Princeton, NJ: ETS.

Bridgeman, B. Morgan, R., and Wang, M. (1997) Choice among essay topics: Impact on performance and validity, JEM, 34, 276-286.

Shaeffer, G. Briel, J., and Fowles, M.E. (2001) Psychometric Evaluation of the New GRE Writing Assessment. GRE Research Report 96-11. Princeton, NJ: ETS.

Willingham, W.W., & Cole, N.S. (1997) Gender and fair assessment. Mahwah, NJ: Erlbaum.

Appendix A

Psychometric Characteristics of the Writing Assessment

Reliability of ratings by Scorer 1 and Scorer 2

Prompt	N	Cronbach Alpha
Issue	8736	0.86
Argument	8736	0.89

Correlation between Issue and Argument Tasks: .57

Correlations among Writing and GRE Analytical, Quantitative and Verbal

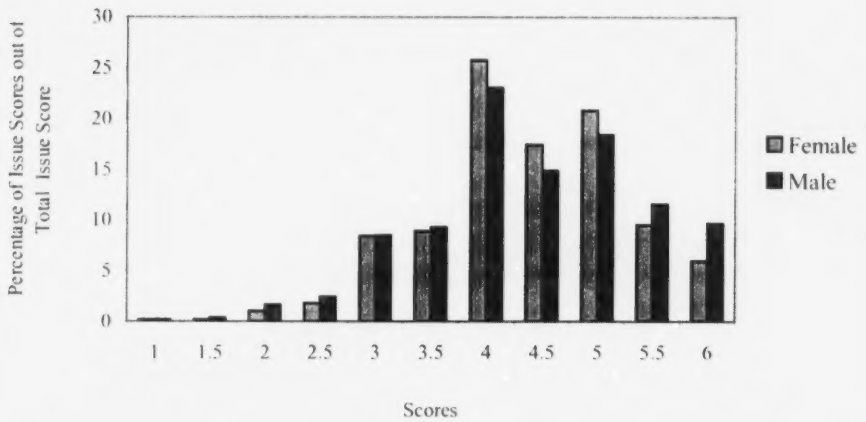
		Pearson Correlation		
Prompt	N	Analytical	Quantitative	Verbal
Issue	5946	0.30	0.21	0.43
Argument	5946	0.35	0.29	0.34
Total score	5946	0.37	0.28	0.49

Appendix B

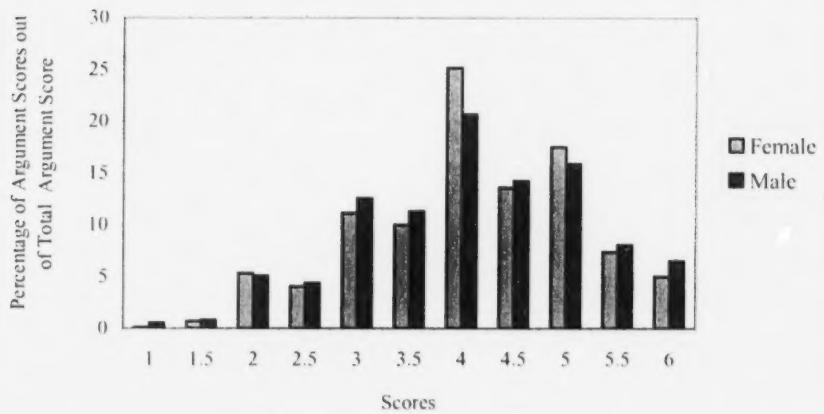
Score Distributions for Examinee Subgroups

Gender Groups

Distribution of Issue Scores by Gender

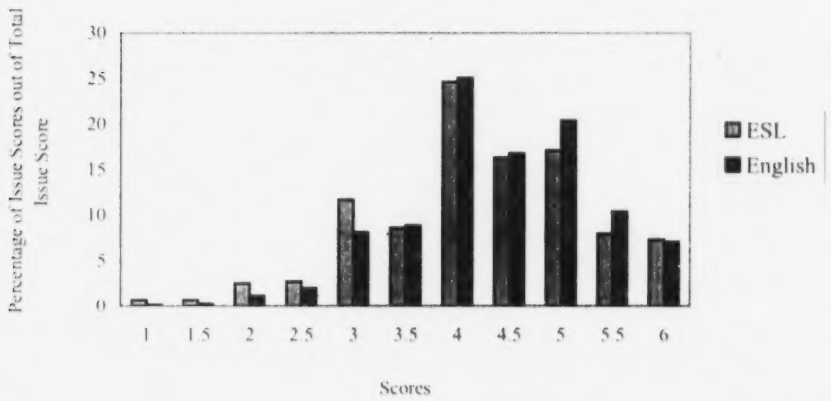


Distribution of Argument Scores by Gender

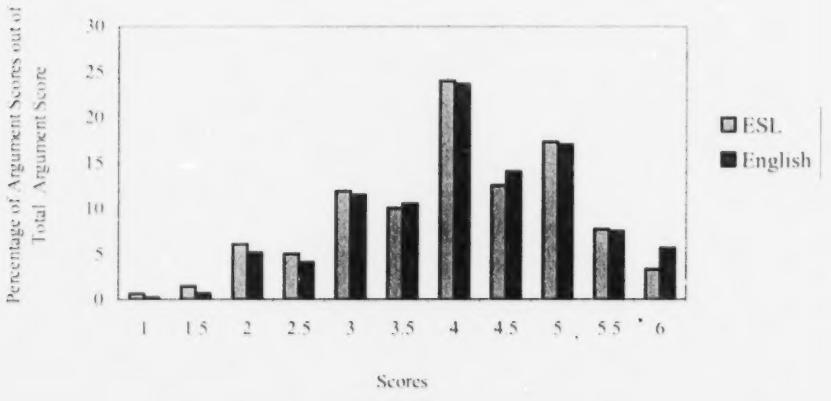


Language Groups

Distribution of Issue Scores by English Best



Distribution of Argument Scores by English Best



Minority Groups

Issue Task

	AI	%	AA	%	MA	%	AsA	%	PR	%	OH	%	W	%	O	%	Total
1	0	0.00	3	0.71	0	0.00	2	0.70	0	0.00	2	1.82	3	0.07	1	0.60	11
1.5	1	3.03	11	2.59	0	0.00	6	2.11	4	7.69	3	2.73	7	0.17	1	0.60	33
2	0	0.00	54	12.74	5	5.05	20	7.04	6	11.54	2	1.82	163	4.02	9	5.39	259
2.5	6	18.18	41	9.67	5	5.05	18	6.34	2	3.85	10	9.09	115	2.84	6	3.59	203
3	4	12.12	88	20.75	13	13.13	44	15.49	9	17.31	18	16.36	394	9.73	18	10.78	588
3.5	1	3.03	59	13.92	13	13.13	26	9.15	8	15.38	18	16.36	405	10.00	12	7.19	542
4	6	18.18	77	18.16	26	26.26	56	19.72	13	25.00	26	23.64	1026	25.33	26	15.57	1256
4.5	7	21.21	31	7.31	18	18.18	32	11.27	5	9.62	9	8.18	596	14.71	32	19.16	730
5	4	12.12	43	10.14	14	14.14	44	15.49	1	1.92	15	13.64	751	18.54	36	21.56	908
5.5	0	0.00	14	3.30	3	3.03	21	7.39	3	5.77	4	3.64	331	8.17	18	10.78	394
6	4	12.12	3	0.71	2	2.02	15	5.28	1	1.92	3	2.73	260	6.42	8	4.79	296
Total	33		424		99		284		52		110		4051		167		5220
Missing=726																	

Argument Task

	AI	%	AA	%	MA	%	AsA	%	PR	%	OH	%	W	%	O	%	Total
1	0	0.00	3	0.71	0	0.00	4	1.41	0	0.00	0	0.00	0	0.00	0	0.00	7
1.5	0	0.00	3	0.71	0	0.00	1	0.35	0	0.00	1	0.91	3	0.07	1	0.60	9
2	0	0.00	13	3.07	0	0.00	6	2.11	5	9.62	2	1.82	28	0.69	2	1.20	56
2.5	1	3.03	17	4.01	2	2.02	13	4.58	4	7.69	2	1.82	44	1.09	3	1.80	86
3	5	15.15	73	17.22	12	12.12	34	11.97	11	21.15	10	9.09	256	6.32	11	6.59	412
3.5	0	0.00	58	13.68	13	13.13	40	14.08	7	13.46	11	10.00	309	7.63	14	8.38	452
4	5	15.15	116	27.36	27	27.27	75	26.41	10	19.23	35	31.82	1019	25.15	36	21.56	1323
4.5	6	18.18	57	13.44	19	19.19	32	11.27	5	9.62	17	15.45	717	17.70	37	22.16	890
5	10	30.30	59	13.92	14	14.14	53	18.66	4	7.69	17	15.45	895	22.09	30	17.96	1082
5.5	4	12.12	16	3.77	7	7.07	18	6.34	1	1.92	12	10.91	458	11.31	18	10.78	534
6	2	6.06	9	2.12	5	5.05	8	2.82	5	9.62	3	2.73	322	7.95	15	8.98	369
Total	**33**		**424**		**99**		**284**		**52**		**110**		**4051**		**167**		**5220**
Missing=726																	

Key:
AI = American Indian
AA = African American
MA = Mexican American
AsA = Asian American, Pacific Islander
PR = Puerto Rican
OH = Other Hispanic
W = White
O = Other

BEST COPY AVAILABLE

Appendix C

Standardized Differences for Examinee Subgroups

on Writing Assessment and General Test

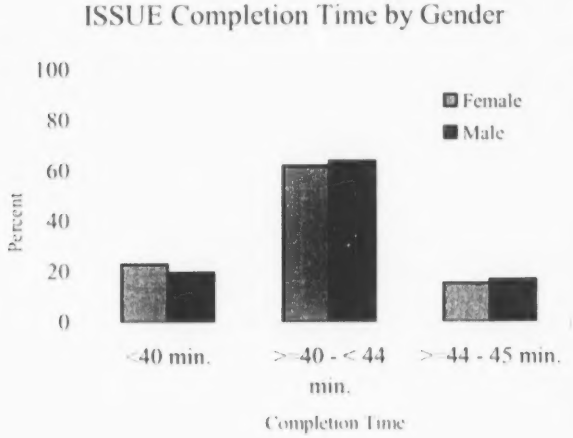
Group	N	Standardized Difference[9]			
			GRE General Test		
		W	V	Q	A
Gender groups:					
Females	4053				
Males	1736	.01	.22	.42	.05
Language groups:					
ESL	479				
Non ESL	5169	.17	.11	-.14	.08
Ethnic groups:					
American Indian	33	0.14	0.24	0.31	0.26
African American	424	0.85	0.89	1.07	1.09
Mexican American	99	0.34	0.59	0.54	0.66
Asian American/ Pacific Isl.	284	0.49	0.43	-0.32	0.29
Puerto Rican	52	0.93	0.58	0.49	0.73
Other Hispanic	110	0.47	0.43	0.39	0.51
Other	167	0.02	-0.08	-0.10	0.03
White (non Hispanic)	4051				
Major field:					
Business	187	.53	.72	-.02	.43
Education	403	.70	.03	.49	.53
Engineering	132	.62	.27	-1.41	-.39
Natural Sciences	1899	.45	.64	-.37	-.05
Social Sciences	1191	.21	.57	.07	.16
Other	651	.30	.53	.05	.16
Humanities and Arts	4273				

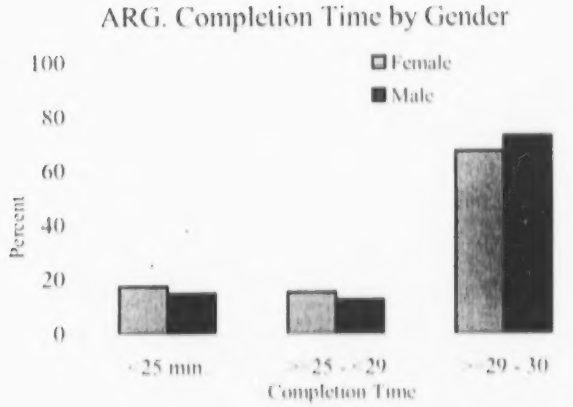
[9] The standardized difference is computed by subtracting the female mean score from the male mean over the pooled standard deviation and by subtracting the English-not-best-language from the English-best-language mean over the pooled standard deviation. For the ethnic groups, each ethnic group is subtracted from the White mean. For the Major field groups, each major field group was subtracted from the Humanities and Arts group.

Appendix D

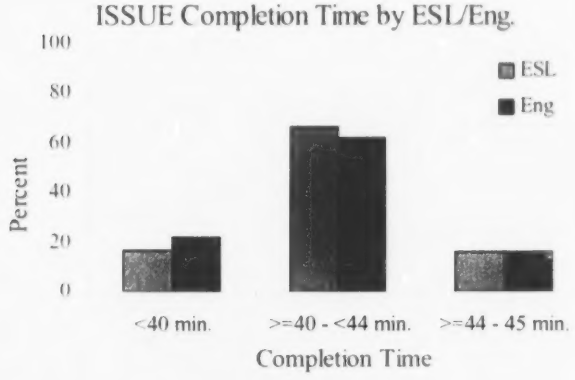
Distributions for Completion Times

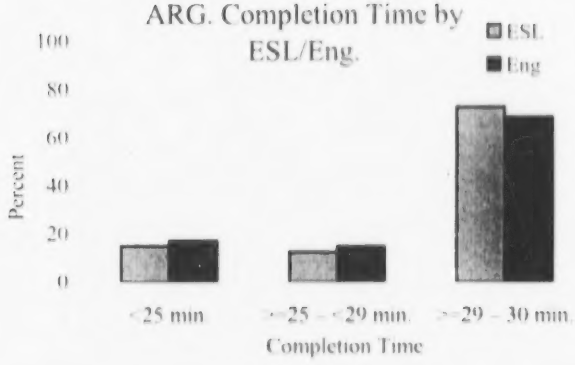
Distributions for Gender Groups





Distributions for Language Groups





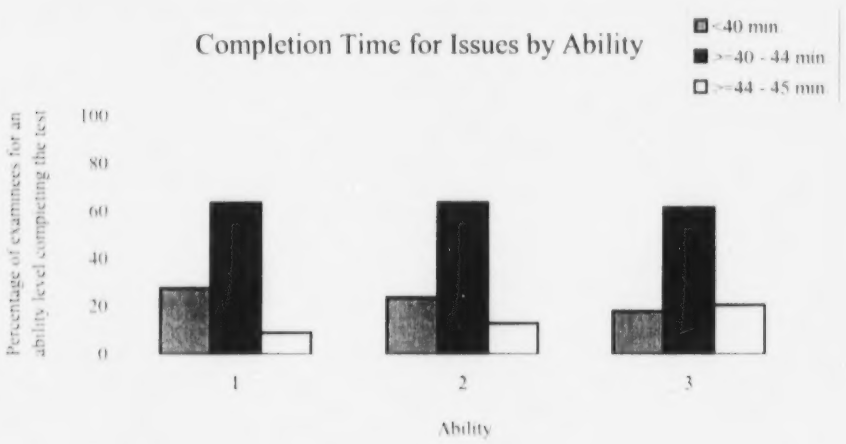
Distributions for Minority Groups

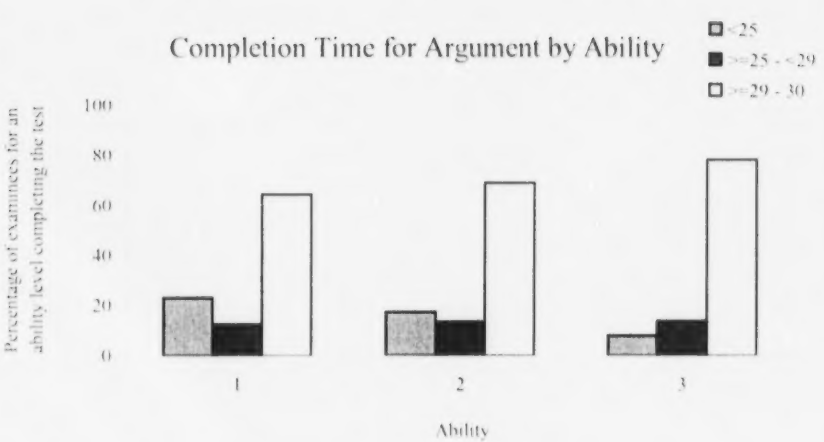
Group	N	% Completing Issue Task		
		< 40 min	40 to < 44 min	44 – 45 min
American Indian	33	12	67	21
African American	424	26	59	15
Mexican American	99	20	72	8
Asian American	284	14	66	20
Puerto Rican	52	17	72	12
Other Hispanic	110	25	63	13
White	4051	23	62	15
Other	167	21	60	19

Group	N	% Completing Argument Task		
		< 25 min	25 – < 29 min	29 – 30 min
American Indian	33	12	18	70
African American	424	21	13	66
Mexican American	99	16	9	75
Asian American	284	7	8	85
Puerto Rican	52	17	13	69
Other Hispanic	110	21	12	67
White	4051	18	16	66
Other	167	15	13	72

Distributions for Ability Groups

In the figures below, examinees were grouped by total score. The low ability group includes scores of 1 – 2.5; middle ability includes scores of 3 – 4.5; and high ability includes scores of 5 - 6.





28

Appendix E

Chi Square Results for Completion Times

Gender Groups

Note: In this examinee population, female examinees far outnumber male examinees.

Issue Topic Chi Square:	8.47 (df = 2)	p = .01
Argument Topic Chi Square:	17.70 (df = 2)	p = .0001

Language Groups

Note: In this examinee population, English-best language examinees far outnumber ESL examinees

Issue Topic Chi Square:	6.85 (df = 2)	p = .03
Argument Topic Chi Square:	3.89 (df = 2)	p = .14

Minority Groups

Note: In this examinee population, White examinees far outnumber other subgroups.

Issue Topic Chi Square:	28.73 (df = 14)	p = .01
Argument Topic Chi Square:	54.25 (df = 14)	p < .0001

TM033804

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

Reproduction Release

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Performance of Examinee Groups on a Measure of Analytical Writing

Author(s): Kathleen O'Neill and Saba Rizavi

Corporate Source: PAPER PRESENTED IN THE ANNUAL MEETING OF THE NATIONAL COUNCIL OF MEASUREMENT IN EDUCATION, NEW ORLEANS, LA

Publication Date: APRIL, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY SAMPLE TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	**Level 2A**	**Level 2B**
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Kathleen O'Neill

Printed Name/Position/Title: KATHLEEN O'NEILL/ PROGRAM ADMINITRATOR

Organization/Address: KATHLEEN O'NEILL
EDUCATIONAL TESTING SERVICE,06-L
PRINCETON, NJ 08541

Telephone: 609-683-2988

Fax:

E-mail Address: KONEILL@ETS.ORG

Date:

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Assessment and Evaluation **1129 Shriver Laboratory (Bldg 075)** **College Park, Maryland 20742**	**Telephone: 301-405-7449** **Toll Free: 800-464-3742** **Fax: 301-405-8134** **ericae@ericae.net** **http://ericae.net**

EFF-088 (Rev. 9/97)